基于地物高光谱和无人机多光谱的黄河三角洲土壤盐分机器学习反演模型

范承志[1]，王梓文[1]，杨兴超[1]，罗永开[2]，徐学欣[3]，郭　斌[1]，李振海[1*]

（1. 山东科技大学 测绘与空间信息学院，山东青岛 266590；2. 滨州学院山东省黄河三角洲生态环境重点实验室，山东滨州 256603；3. 青岛农业大学 农学院，山东青岛 266109）

摘　要：土壤盐渍化是限制黄河三角洲地区农业经济发展的重要因素，进一步阻碍了农业生产。为了探索无人机影像在地表无植被覆盖条件下的土壤盐分含量反演状况，以黄河三角洲典型区域为研究区，获取地物高光谱和无人机多光谱两种数据源与样点土壤盐分含量，通过优选敏感光谱参量，使用偏最小二乘回归（Partial Least Squares Regression，PLSR）和随机森林（Random Forest，RF）两种机器学习算法建立土壤盐分含量反演模型，实现研究区的土壤盐分含量反演。结果表明：（1）高光谱1972 nm波段与土壤盐分含量间的敏感性最高，相关系数为−0.31。（2）两种不同数据源优化后的RF模型均优于PLSR，且稳定性更好。（3）基于地物高光谱的RF模型（R^2=0.54，RMSEv＝3.30 g/kg）优于基于无人机多光谱的RF模型（R^2=0.54，验证RMSRv＝3.35 g/kg）。（4）结合无人机影像采用多光谱RF模型对研究区耕地的土壤盐分含量进行反演，研究区总体以轻、中度盐渍化土壤为主，对作物的耕种具有一定程度的限制。本研究构建并对比了两种不同源数据的黄河三角洲土壤盐分反演模型，并结合各自数据源的优势进行优化，探索了地表无植被覆盖情况下的土壤盐分含量反演方法，对更精准反演土地盐渍化程度提供了参考。

关键词：土壤盐分含量；遥感；地物高光谱；无人机多光谱；偏最小二乘回归；随机森林；机器学习

中图分类号：S127　　**文献标志码：**A　　**文章编号：**SA202212001

引用格式：范承志，王梓文，杨兴超，罗永开，徐学欣，郭斌，李振海．基于地物高光谱和无人机多光谱的黄河三角洲土壤盐分机器学习反演模型[J]. 智慧农业(中英文), 2022, 4(4): 61-73.

FAN Chengzhi, WANG Ziwen, YANG Xingchao, LUO Yongkai, XU Xuexin, GUO Bin, LI Zhenhai. Machine learning inversion model of soil salinity in the yellow river delta based on field hyperspectral and UAV multispectral data[J]. Smart Agriculture, 2022, 4(4): 61-73. (in Chinese with English abstract)

1　引　言

土壤盐渍化是世界范围内重要的资源及生态问题，已经成为制约区域农业生产和经济可持续发展的主要障碍之一[1,2]。受盐渍化影响的土壤有机质含量较少、肥力低、理化性质差，作物不易促苗，在重度盐碱化的土壤中植物很难生存。改善传统的土壤盐度测量方法，大范围、快速且准确地监测耕地土壤的盐渍化水平成为提高农业经济健康发展的重要前提。黄河三角洲地区是中国典型的土壤盐渍化区域，该地区由于黄河水侧渗、海水倒灌以及水土资源的不合理利用等因

收稿日期：2022-12-13

基金项目：山东省自然科学基金（ZR2022MD017）；河北省重点研发计划项目（22326406D）

作者简介：范承志（1998—），男，硕士，研究方向为农业定量遥感。E-mail：fancz98@163.com

*通信作者：李振海（1989—），男，博士，教授，研究方向为农业遥感。E-mail：lizh323@126.com

素，造成了土壤中次生盐渍化严重。土壤退化、板结，以及肥力降低等问题极大地限制了农作物的生长，严重阻碍了黄河三角洲地区农业经济的可持续性发展[3]。

传统的土壤盐度测量方法主要通过实地采样和实验室分析实现，费时费力，很难对农用地进行大面积全覆盖的动态监测[4]。随着遥感技术的快速发展，通过遥感技术对土壤盐分含量进行准确地定量反演成为了许多研究者关注的重点[5]。在多源影像方面，Wang等[6]通过构建Cubist模型对比Landsat-8 OLI（Operational Land Imager）与Sentinel-2 MSI（Multi-spectral Instrument）传感器之间的差异，发现更高空间分辨率和光谱分辨率的遥感影像有助于提高模型反演的准确性；孙亚楠等[7]通过对卫星尺度的Landsat-8 OLI和实测高光谱数据进行融合，构建内蒙古河套灌区地区的高-多光谱融合模型，表明高光谱和多光谱数据融合能够提高卫星遥感反演土壤盐分含量的精度。Hu等[8]使用电磁感应（Electromagnetic Induction，EMI）设备和高光谱数据构建新模型，与卫星影像构建的土壤盐分反演模型作对比，表明无人机（Unmanned Aerial Vehicle，UAV）多光谱数据建模结果更优。搭载多光谱传感器的无人机在农业遥感监测方面兼具多方面优势，与卫星遥感技术相比，具有高空间分辨率、高灵活度的特点；与地面高光谱数据相比，具有图谱合一、方便快捷的优势。在地表植被长势与土壤盐分含量间的联系方面，奚雪等[9]构建无人机-卫星一体化的偏最小二乘法土壤盐分反演模型，在黄河三角洲地区的麦区构建的绿度指数与土壤盐渍化水平之间建立联系，以反演土壤盐分含量。贾吉超等[10]研究了冬小麦种植区与土壤盐分含量间的联系，发现冬小麦种植范围及其变化显著受土壤盐分状况及其变化的影响，土壤含盐量升高冬小麦种植面积骤减。黄权中等[11]在河套灌区进行定点监测并结合OLI影像，研究发现土壤含盐量与玉米、向日葵的长势和产量呈明显负相关。在传统光谱指数优化方面，杨宁等[12]引入红边波段构建新的光谱指数，结果表明红边植被指数模型对内蒙古灌区的土壤盐分含量反演效果更优。多光谱影像包含地物的光谱信息和图像信息，纹理特征是图像空间灰度信息的重要属性。目前，纹理特征为重要的影像分类依据。黄静等[13]利用Sentinel-2影像结合纹理特征显著提升对黄河三角洲地区中度盐渍土的分类效果，有效提高影像分类和盐渍化信息监测精度。台翔[14]通过分析不同盐分条件下土壤和植被的光谱特征和图像纹理，构建了植被覆盖条件下的土壤盐分反演模型，加入纹理特征参数能有效提高模型精度。

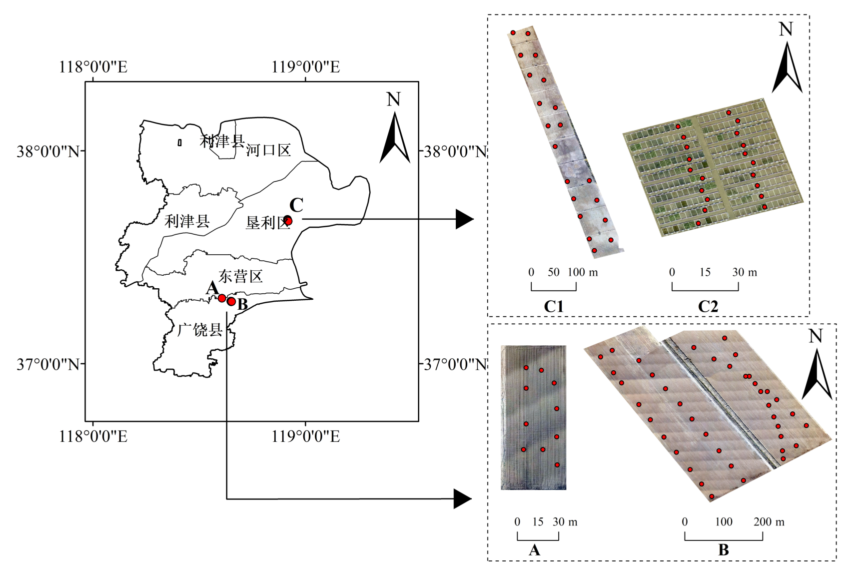
目前通过遥感技术反演土壤含盐量的研究存在以下两个问题。一是大多研究集中在近红外波段附近，已有GF-5、HJ-1A等卫星携带了高光谱传感器，但受限于空间分辨率和波段数目，难以在较小尺度内精确预测土壤盐分含量；二是随着无人机遥感的广泛应用，纹理特征的引入能否提高土壤盐分反演精度。因此，本研究通过实地采样调研，将实测的地物高光谱与无人机多光谱的地表反射率计算多个光谱指数并进行优选，分别建立两种数据源的土壤盐分反演模型，并结合各自数据特征进行优化，对比分析两种数据源的反演模型结果并评价，最终使用无人机影像结合精度较好的模型反演研究区的土壤盐渍化水平。

2 数据与方法

2.1 研究区概况

研究区位于山东省东营市（118°07′—119°10′E，36°55′—38°10′N），是黄河三角洲地区典型的土壤盐渍化区域，属于暖温带半湿润大陆性季风气候，年均降水量少但蒸散量较大，主要土壤类型为潮土和盐土，地势平坦、土壤盐渍化特征明显[9]，符合研究需要。综合东营市区域调查结果、地貌土壤以及农作物的分布情况，分别在研究区的南部和东北部的农作物种植区选择了A、B、C三个小研究区，其中研究区A、B主

要种植作物是小麦，研究区C分为距离相近C1和C2两部分，C1种植水稻，C2为含盐量差异性明显的裸土（图1）。研究区整体分布上具有典型性和空间异质性，能够体现黄河三角洲地区盐渍化土壤的分布趋势。



注：图中A、B、C为研究区

图1 山东省东营市黄河三角洲试验研究区分布图

Fig. 1 Distribution map of Yellow River Delta experimental study area in Dongying city, Shandong province

2.2 地面数据采集与处理

于2021年11月23日至25日在研究区内进行了实地调查，发现该时间段研究区降水量较少，地表特征稳定，土壤盐渍化较为明显，便于提取土壤端元的光谱特征。为确保均匀随机取点，在研究区内一共采集90个样本点，其中研究区A样本10个，研究区B样本40个，研究区C1样本20个，研究区C2样本20个。

使用华星A12系列全球导航卫星系统（Global Navigation Satellite System，GNSS）接收机，测量每个样本点的WGS84大地坐标。待测得土壤高光谱数据后，采集0～5 cm深度的土壤作为待测盐度样本。每个样本称取20 g的烘干土壤加入去离子水，搅拌、抽滤得到土壤样本浸出液。而后将土壤样本浸出液倒入干燥至恒重的蒸发皿中进行蒸干，加入过氧化氢（30%）去除有机质后，放入烘箱中烘干至恒重，冷却后称量，从而确定每个样本的土壤含盐量（Soil Salinity，SS，g/kg）[15]。

2.3 光谱数据采集及预处理

2.3.1 地物高光谱数据采集

在研究区使用SVC HR-768i型便携式地物光谱仪（Spectra Vista Corporation，纽约，美国）测量样本点表层土壤端元的光谱特征。SVC HR-768i型地物光谱仪共有768个通道用以接收波长在350～2500 nm之间的电磁辐射数据，在350～1000 nm间的通道带宽小于1.5 nm，在1000～2500 nm之间带宽为4～8 nm。观测时天气晴朗，传感器探头距离地面1 m，垂直正射样点土体自然状态的表面，每次测量前均使用标准白板校准反射率，每次光谱扫描时间设置为5 s，同时对每个样点重复测量5次，取反射率均值作为样点的土壤高光谱反射率曲线。

2.3.2 无人机多光谱数据采集

使用大疆DJI精灵4Pro多光谱版无人机采集地面多光谱数据，摄像头带有1个可见光成像的

彩色传感器和5个多光谱成像的单色传感器：蓝（B）、绿（G）、红（R）、红边（RE）、近红（NIR）光谱段，中心波长分别为450、560、650、730和840 nm。多光谱传感器通道宽于高光谱，近红外波段通道宽度为±26 nm，其他波段均为±16 nm。飞行前，将标准灰板放置在研究区内对传感器进行定标，试验期间无人机飞行时间为10:00—14:00，飞行高度100 m，影像采集间隔2 s，晴朗无风。

使用DJI Terra软件对无人机多幅影像进行拼接并设置地理参考，得到研究区的各波段及RGB影像。通过标准灰板的各项参数对多波段影像进行辐射定标，将灰度级转换为光谱反射率。在ENVI 5.3中进行无人机影像的波段融合和裁剪，结合实地测量样本点的GPS（Global Positioning System）坐标数据在ArcGIS 10.6中提取样本坐标点对应像元的5个波段反射率数据。

2.4 光谱指数的选择及纹理特征的引入

2.4.1 光谱指数

依照前人研究进展，计算多个光谱指数，如表1所示。针对无人机多光谱传感器获取的红边波段影像，引入使用红边波段构建的新型光谱指数[10]。

表1 光谱指数计算公式

Table 1 Formula of spectral index

光谱指数	计算公式		参考文献
归一化植被指数(Normalized Difference Vegetation Index，NDVI)	$NDVI = (NIR - R)/(NIR + R)$	(1)	[16]
土壤调节植被指数(Soil-Adjusted Vegetation Index，SAVI)	$SAVI = 1.5 \times (NIR - R)/(NIR + R + 0.5)$	(2)	[17]
改进非线性指数(Modified Nonlinear Index，MNLI)	$MNLI = 1.5 \times (NIR^2 - R)/(NIR^2 + R + 0.5)$	(3)	[18]
光谱指数SAVI*SR	$SAVI*SR = (NIR^2 - R)/[(NIR + R + 0.5) \times R]$	(4)	[18]
改进叶绿素吸收比值指数(Modified Chlorophyll Absorption In Reflectance Index1，MCARI1)	$MCARI1 = [(NIR - RE) - 0.2 \times (NIR - R)](NIR/RE)$	(5)	[19]
改进红边指数(Modified Red Edge Simple Ratio，MRESR)	$MRESR = (NIR - R)/(RE - R)$	(6)	[20]
标准化红边指数(Normalized Red Edge Index，NREI)	$NREI = RE/(NIR + RE + R)$	(7)	[20]
比值植被指数(Ratio Vegetation Index，RVI)	$RVI = NIR/R$	(8)	[21]
最优植被指数(Optimal Vegetation Index，VI_{opt})	$VI_{opt} = 1.45 \times ((NIR^2 + 1)/(R + 0.45))$	(9)	[22]
归一化盐分指数(Normalized Differential Salinity Index，NDSI)	$NDSI = (R - NIR)/(R + NIR)$	(10)	[23]
盐分指数SI1(Salinity Index)	$SI1 = SQRT(G \times R)$	(11)	[23]
盐分指数SI3(Salinity Index)	$SI3 = SQRT(G^2 + R^2)$	(12)	[23]
盐分指数SI1-reg(Salinity Index)	$SI1reg = SQRT(G \times RE)$	(13)	[12]
盐分指数SI3-reg(Salinity Index)	$SI3reg = SQRT(G^2 + RE^2)$	(14)	[12]
红边归一化盐分指数(Red-Edge Based Normalized Differential Salinity Index，NDSI-reg)	$NDSIreg = (RE - NIR)/(RE + NIR)$	(15)	[12]

注：NIR、RE、R、G、B分别代表近红外、红边、红、绿、蓝波段，在高光谱数据中分别代表841、730、650、560和450 nm波段；reg代表计算中用红边波段替换红波段

2.4.2 纹理特征

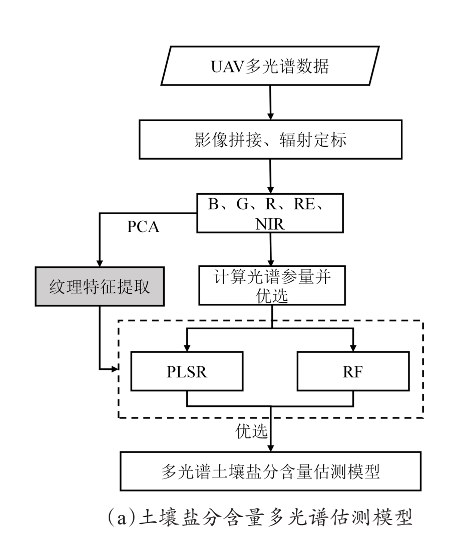
影像纹理反映了土壤表面颜色和灰度的变化，这些变化与土壤本身的盐度、含水量等因素紧密相关。本研究采用基于统计的灰度共生矩阵法提取影像纹理特征。灰度共生矩阵（Gray Level Co-Occurrence Matrix，GLCM）是一种在特征提取、纹理分析以及影像质量评价方面广泛应用的重要方法，用以描述图像中像素灰度之间的相关性[24]。由于灰度共生矩阵的维度较大，一般不直接作为区分纹理的特征，而是基于矩阵构建的统计量作为描述图像纹理的特征。

在ENVI 5.3中对原始的无人机多波段影像进行主成分分析（Principal Component Analysis，PCA），对多波段影像进行降维。无人机影像的第一主成分的贡献率达95%以上，故提取影像的第一主成分，进行基于灰度共生矩阵的影像纹理

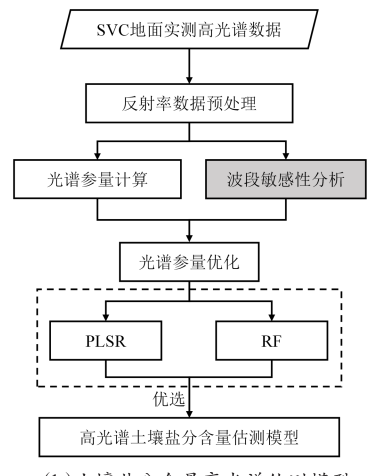
特征提取，共得到8个影像的纹理特征，包括均值（Mean）、方差（Variance）、同质性（Homogeneity）、对比度（Contrast）、差异性（Dissimilarity）、熵（Entroy）、角二阶矩（Angular Second Moment，ASM）和相关性（Correlation）[25]。

2.5 土壤盐分含量反演模型的构建与验证

从地面高光谱和无人机多光谱两方面分别构建土壤盐分含量模型，思路如图2所示。



（a）土壤盐分含量多光谱估测模型



（b）土壤盐分含量高光谱估测模型

图2 土壤盐分含量估测技术流程图

Fig.2 Technical flow chart of soil salt content estimation

（1）样本划分。将90个样本随机分成60个训练样本和30个验证样本，对60个训练样本的土壤含盐量数据与光谱反射率数据计算的多个光谱指数进行Pearson相关性分析。

（2）光谱及纹理特征筛选。无人机多光谱数据以纹理特征和光谱指数进行筛选；地面高光谱在原始输入波段基础上，结合敏感波段进行进一步优化。根据土壤的光谱特征，使用无人机采集的蓝（B）、绿（G）、红（R）、红边（RE）、近红外（NIR）共5个波段的数据计算15个植被指数和盐分指数（表1）。SVC地物高光谱仪能够采集350~2500 nm范围内多个窄光谱段的反射率数据，但是与无人机多光谱相机传感器的中心波长并不完全重合。为保持一致性，选择与多光谱传感器中心波长尽可能相近的光谱段。选择实测的土壤高光谱反射率数据中450、560、650、730和841 nm通道分别代表无人机影像中的B、G、R、RE、NIR波段。特征优选方法采用相关性分析，排除无法通过共线性诊断或与盐分相关性较低的光谱指数[9]。r表示光谱参量间的相关系数，r超过0.9时认为存在共线性问题，在0.8以上时可能会有问题[26]。

（3）模型选择。由于遥感计算中各光谱指数之间存在比较严重的多重相关性，为减弱模型自变量间多重共线性问题，采用偏最小二乘回归（Partial Least Squares Regression，PLSR）和随机森林（Random Forest，RF）建模。研究需要采用多个光谱指数构建模型，偏最小二乘回归模型集合了主成分分析、线性回归分析和相关分析的优势，在预测变量间存在高度相关甚至共线问题且样本数量较小时，与多元线性回归模型（Maximum Likelihood Estimation，MLR）相比更具优势。RF模型将多个互不联系的决策树构成森林，对多元共线性问题不敏感，结果对样本中的缺失数据和不平衡的数据比较稳健，是机器学习中应用较为广泛的算法[27]。选择与土壤盐分含量显著相关的光谱参量，利用Matlab R2021a中构建土壤盐分含量的估测模型。

（4）精度评价。以土壤含盐量的实测值与预测值之间的决定系数R^2和验证集均方根误差

（Root Mean Squared Error of Validation，RMSEv）综合评价模型的训练和验证精度。

（5）盐度分级及制图。选择较优的一种建模方法建立黄河三角洲土壤盐分含量的遥感反演模型，估测研究区耕地的土壤盐渍化水平。使用ENVI 5.3对无人机多波段影像进行处理，通过目视解译提取影像中耕作区域。综合对比所构建的PLSR模型和RF模型的精度，选择最佳的裸土土壤盐分含量遥感估测模型，用其对研究区内的耕地进行无人机遥感尺度的土壤盐分含量反演。同时对耕地的土壤盐渍化程度进行分级，将土地的盐渍化程度按照国家土壤盐渍化分级标准划分为5个等级（表2），并统计面积占比[9]。

表2　土壤盐渍化程度分级标准

Table 2 Grading standards for soil salinization

土壤盐渍化等级	非盐渍化	轻度盐渍化	中度盐渍化	重度盐渍化	盐土
土壤含盐量/($g \cdot kg^{-1}$)	<1	1～2	2～4	4～6	>6
等级	1	2	3	4	5

3　模型优化与结果分析

3.1　土壤盐分与光谱特征相关性分析

3.1.1　基于地物高光谱的土壤盐分指数分析及优化

对比低盐度与高盐度样本光谱特征差异，分析地面多光谱数据各波段反射率与土壤盐分含量之间相关性（图3）结果表明，对350～2500 nm间全部的光谱段进行Pearson相关性分析，在350～1350 nm间为正相关，1350～2500 nm间为负相关；在1400和1850 nm波段附近存在剧烈变化，1972～1978 nm间的反射率对盐度的敏感性最高，其中1972 nm波段的相关系数r为最敏感的−0.31。可见光波段均未达极显著相关，绿光和蓝光相关性最强，红边波段敏感性较弱r接近0。短波红外（Short Wavelength Infra-Red，SWIR）与近红外波段相较，具有更加敏感的负相关性，考虑使用SWIR构建光谱参量。

分析地面高光谱数据与土壤含盐量相关性

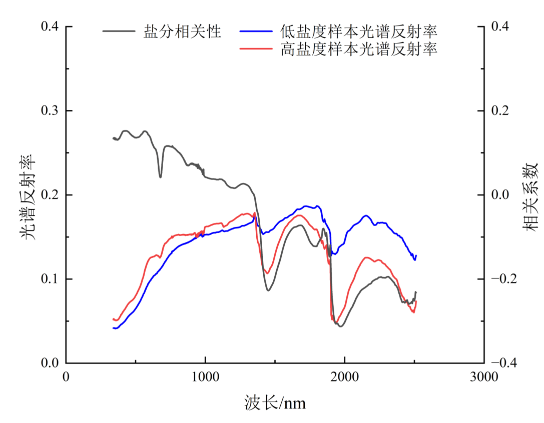


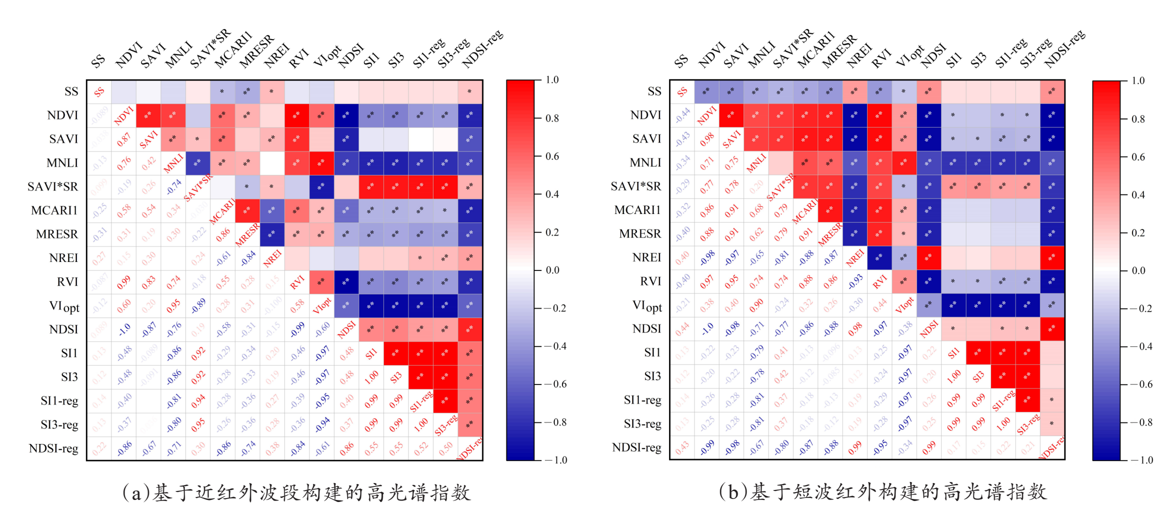
图3　各光谱段与土壤含盐量间的相关性分析

Fig. 3 Correlation analysis between spectral bands and soil salt content

（图4（a）），结果表明，基于近红外波段构建的光谱指数，MCARI1、NREI、MRESR、NDSI-reg达到0.05水平上显著相关；NREI和MRESR两个光谱指数与土壤盐分含量达到显著相关，且彼此间不存在极强的共线性问题，MRESR为正相关，NREI为负相关。通过全光谱段的相关性分析，R和NIR波段的反射率敏感性较差，选择与土壤盐分含量相关性更优的波段重新计算光谱参量以提高模型精度，所以选择使用SWIR 1972 nm代替原始光谱指数中的NIR 841 nm。结果表明，使用SWIR 1972 nm重构光谱参量后，所有包含近红外波段运算的光谱指数与土壤含盐量间的相关性均得到了大幅度提升（图4（b））。除VI_{opt}及SI类指数外，其他光谱参量均达极显著相关。在相关系数的矩阵中可以看出，光谱参量间的共线性问题有所加重，MCARI1与MRESR之间的r为0.91，认定两个指数间存在共线性问题，因此，后续建模中选择MNLI、SAVI*SR、MRESR和NDSI作为构建模型的自变量。

3.1.2　基于无人机多光谱的土壤盐分相关性分析

无人机影像提取地面点反射率数据，与土壤盐分相关性分析，筛选满足0.01水平极显著相关且不存在严重共线性问题的光谱指数（图5）。SAVI*SR、VI_{opt}、SI1、SI3、SI1-reg和SI3-reg与样本含盐量均达到极显著相关。在盐分指数中，



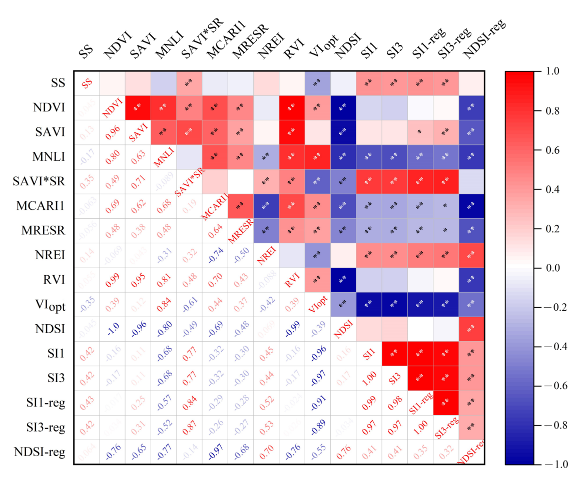
（a）基于近红外波段构建的高光谱指数　　（b）基于短波红外构建的高光谱指数

注：**表示在0.01水平（双侧）上显著相关；*表示在0.05水平（双侧）上显著相关

图4　高光谱指数与土壤盐分相关性分析

Fig. 4 Correlation analysis between hyperspectral indices and soil salinity

除NDSI外，SI1、SI3、SI1-reg、SI3-reg敏感性均表现较好，但是指数间存在极强的共线性问题，相关系数r接近1。相较SWIR，可见光区间内敏感性较低，红边波段不是土壤盐度的敏感波段区间，绿光波段的贡献性更高。VI_{opt}与4个盐分指数均纯在较高的相关性，存在共线性问题，选择VIF<10的VI_{opt}（r = −0.35）、SAVI*SR（r = 0.35）、SI3-reg（r = 0.42）三个光谱指数构建无人机尺度的土壤盐分估测模型。



注：**表示在0.01水平（双侧）上显著相关；*表示在0.05水平（双侧）上显著相关

图5　无人机多光谱指数与土壤盐分相关性分析

Fig. 5 Correlation analysis between multi-spectral indices of UAV and soil salinity

3.2 无人机影像的纹理特征分析

考虑无人机多光谱影像数据的优势，引入影像的纹理特征作为新的模型自变量。对PCA处理后的影像计算，获得8个纹理特征波段后，按照坐标提取各样本点对应的纹理特征并与盐度进行敏感性分析（表3）。无人机影像提取纹理特征会造成的数据冗余，对8个纹理特征两两之间计算相关系数r，排除呈高度相关的特征变量，减弱数据冗余对模型精度的影响。同质性（r = 0.16）、对比度（r = −0.15）和差异性（r = −0.16）3个特征向量的敏感性接近，在无人机图像中各样本点的区分度不高，对盐度的敏感性有限。由于纹理特征受到图像噪声的一定影响，在8个纹理特征中，土壤含盐量仅与熵（r = −0.26）、角二阶矩（r = 0.27）两个特征向量满足在0.05水平上显著相关，其他特征向量的相关性均表现不高，对模型精度的提高没有意义。但是，熵和角二阶矩二者间存在较高的一致性，r为0.97，认定二者间存在严重共线问题，存在严重共线性问题的纹理特征对于土壤盐度反演模型的构建是没有帮助的，因此本研究选择角二阶矩与其他光谱参量作为构建无人机多光谱反演模型的自变量。

表3 无人机影像各纹理特征与土壤盐分含量相关性

Table 3 Correlation between texture characteristics and soil salt content

	SS	Mean	Variance	Homogeneity	Contrast	Dissimilarity	Entroy	ASM	Correlation
SS	1								
Mean	0.19	1							
Variance	−0.18	0.26*	1						
Homogeneity	0.16	−0.44**	−0.60**	1					
Contrast	−0.15	0.33**	0.83**	−0.81**	1				
Dissimilarity	−0.16	0.40**	0.74**	−0.96**	0.94**	1			
Entroy	−0.26*	0.23*	0.52**	−0.74**	0.52**	0.67**	1		
ASM	0.27*	−0.19	−0.43**	0.69**	−0.43**	−0.59**	−0.97**	1	
Correlation	−0.06	−0.09	0.19	0.04	0.06	0.01	0.14	−0.09	1

注：** 表示在 0.01 水平（双侧）上显著相关；*表示在0.05水平（双侧）上显著相关

3.3 基于地面高光谱数据的土壤盐分估测模型的优化与验证

使用高光谱波段优化后的4个光谱指数作为模型的自变量，分别构建PLSR和RF土壤盐分估测模型（图6）。通过优化前后的模型精度对比，PLSR模型的RMSEv从3.59 g/kg减少到3.38 g/kg，R^2从0.11提升到0.20；RF模型的RMSEv从4.30 g/kg减少到3.30 g/kg，提升更加明显，有较好的精度，R^2从0.37提升到0.54。可以看出SWIR对裸土土壤盐分含量反演精度的提升是有效的，土壤中盐分含量的差异在短波红外波段的相关性变化更具特征性。从图6可以看出，改进后的两种模型对裸土土壤盐分含量的反演能力依旧有限，RF模型的精度要优于PLSR模型，RF模型对土壤盐分反演的效果更优，但是受到土壤样本盐分含量分布不均匀的限制，对盐分含量过高的土壤样本很难做到准确地反演，并且对一些低盐度土壤的反演存在一定高估，但是对大部分中、低盐度样本拟合趋势明显，对高盐度样本的预测结果也有一定程度改善，具有更好的拟合精度。

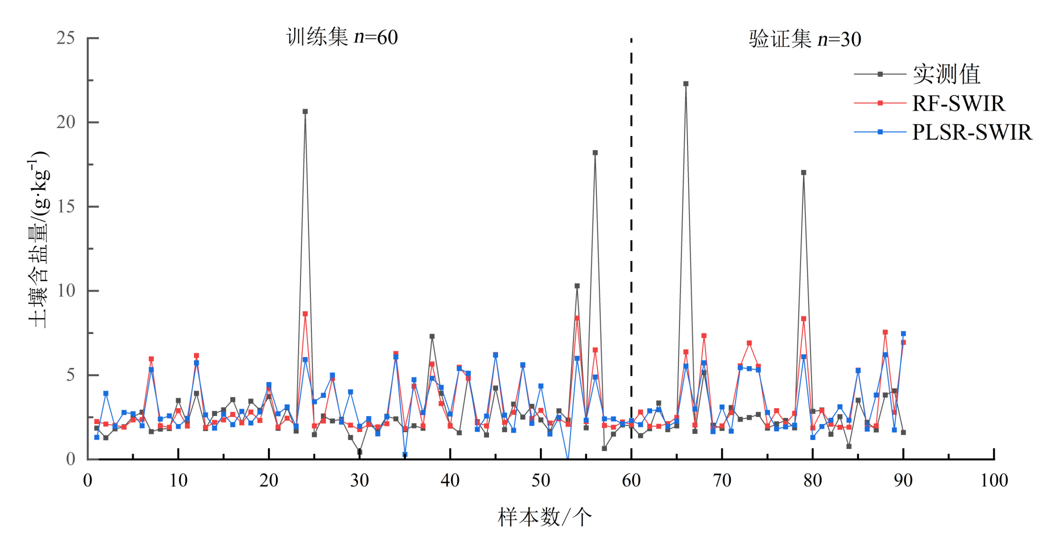


图6 波段优化后偏最小二乘回归模型(PLSR-SWIR)和随机森林模型(RF-SWIR)土壤盐分估测结果

Fig. 6 Soil salinity estimation results of band optimized spectral indices for PLSR model(PLSR-SWIR) and RF model(RF-SWIR)

3.4 基于无人机多光谱影像的土壤盐分反演模型的构建和验证

将角二阶矩作为自变量带入土壤盐分反演模型，与无纹理特征的模型相比（表4），模型的R^2和RMSEv都有所提升。结合无人机纹理特征PLSR模型的R^2为0.26，RMSEv为3.59 g/kg，RF模型R^2为0.54，RMSEv为3.35 g/kg，表明影像纹理特征有助于提升无人机遥感反演土壤盐分含量的精度。相较地面高光谱数据模型（RMSEv=3.30 g/kg），高光谱模型更为稳定，拟合的效果更好。同样，多光谱RF模型的估测效果优于PLSR模型，对估测高盐度的样本准确性有了一定提升，可以看出对验证集预测效果有了一定提升（图7）。受到土壤盐度样本和算法模型自身限制，不能对超越样本范围的数据进行预测，也难以对盐分含量过高和过低的土壤样本准确反演。对高盐度土壤的反演准确度低，是造成模型整体精度较差的主要原因，但是盐度大于15 g/kg样本的反演结果符合盐土的评价标准，在无人机影像制图中影响较小。

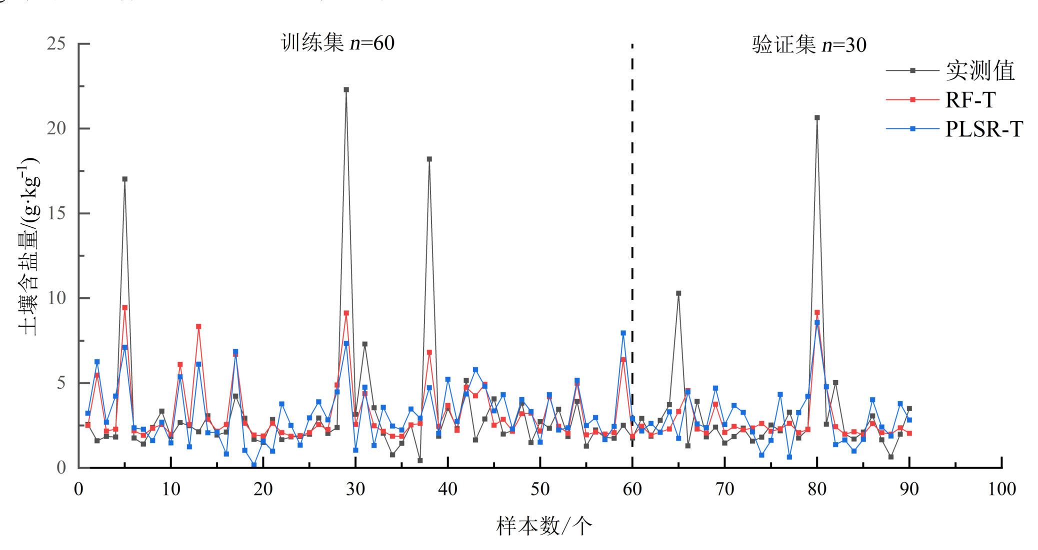


图7 结合纹理特征与无人机多光谱指数的偏最小二乘回归模型(PLSR-T)和随机森林模型(RF-T)土壤盐分估测结果

Fig. 7 Soil salinity estimation results of PLSR(PLSR-T)and RF(RF-T)models combining texture features and UAV based spectral indices

3.5 基于无人机多光谱影像的研究区土壤盐分含量反演

通过PLSR和RF模型的对比，选择使用无人机多光谱数据构建估测土壤盐分的RF模型。按照土壤盐渍化分级标准将研究区内的土壤盐分含量划分为5个等级（表2），统计不同盐渍化等级土壤的面积所占总面积的百分比，评价耕地质量。从研究区内各等级盐渍化土壤的面积统计情况（表5）可以看出的研究区C与A、B相较，整体的土壤盐渍化问题更加严重，土壤盐渍化程度最高，与实地调查情况相符，重度盐渍化和盐土的总面积超过了30%。研究区A、B均为轻、中度盐渍化耕地，研究区C为中、重度盐渍化。结合实际地理空间的分布，符合越靠近黄河入海口，土壤盐渍化问题更严峻的事实。由于土壤盐渍化严重影响作物生长，研究区A、B耕种的主要作物是小麦，而研究区C耕作的是水稻。

4 讨 论

在基于遥感技术对区域土壤盐分含量定量反演的研究中，对地表裸土盐分含量的准确分析相对匮乏，仍然需要深入剖析和探索。本研究使用SVC地物光谱测量仪获取地面高光谱数据，对样本点进行多次测量取均值，保证土壤高光谱数据

表4 土壤盐分含量反演模型对比

Table 4 Inversion model of soil salinity content

数据源	建模方法	精度评价	
		R^2	RMSEv/(g·kg^{-1})
SVC高光谱	PLSR-NIR	0.11	3.59
	PLSR-SWIR	0.20	3.38
	RF-NIR	0.37	4.30
	RF-SWIR	0.54	3.30
UAV多光谱	PLSR	0.19	3.71
	PLSR-T	0.26	3.59
	RF	0.48	3.50
	RF-T	0.54	3.35

注：PLSR-NIR、RF-NIR为近红外高光谱模型，PLSR-SWIR、RF-SWIR为短波红外高光谱模型；PLSR、RF为未引入纹理特征的多光谱模型，PLSR-T、RF-T为结合纹理特征的多光谱模型

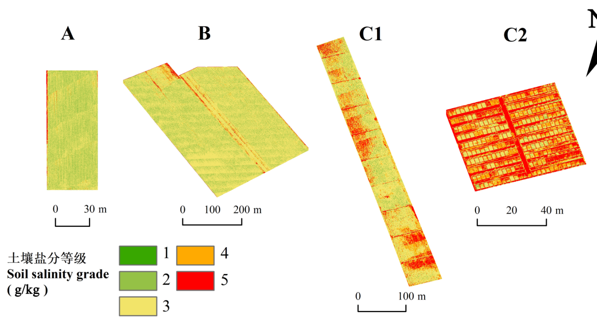


图8 研究区土壤盐分反演图

Fig. 8 Inversion map of soil salinity in the study area

表5 黄河三角洲研究区土壤盐渍化等级面积占比统计

Table 5 Proportion statistics of soil salinization grade area in the study area

单位：%

研究区	等级				
	1	2	3	4	5
A	0	43.1	54.2	1.5	1.2
B	0	25.9	70.2	2.6	1.3
C1	0	10.4	58.7	19.4	11.5
C2	0	4.3	26.1	28.6	41.0

特征的准确提取。使用无人机获取多光谱数据，对研究区先进行无人机测量，后进行实地高光谱测量，保证无人机影像的纹理特征不受人为因素的干扰，尽可能保持土壤的自然状态，保证反演结果更加准确。

引入短波红外波段后的光谱指数与土壤盐分含量的相关性在波长1900 nm附近显著提升，这与前人研究[28]中盐碱土光谱反射率在该波段附近的陡降相吻合。在土壤盐渍化过程中产生NaCl等化合物，这些化学物质与水在键态结构上的结合会产生特异性的光谱吸收峰特征[29]。在1400～1500 nm、1800～1900 nm形成敏感性较强的光谱段，对土壤盐分含量间的差异分析具有更多深入探索的可能。

本研究通过构建高光谱和多光谱两种数据类型的裸土土壤盐分含量估测模型，并进行优化和分析。优化后，两种数据源的PLSR模型分别提升验证集RMSEv分别优化到3.38 g/kg和3.59 g/kg，RF模型的验证RMSEv分别优化到3.30 g/kg和3.35 g/kg，表明无人机影像结合图谱信息能够充分发挥自身优势。RF在泛化能力和平衡误差方面表现良好，但受限于机器学习的鲁棒性，存在一定的过拟合问题[30]。基于地物高光谱数据的RF反演模型（R^2=0.54，验证RMSEv＝3.30 g/kg）与基于无人机多光谱的RF模型（R^2＝0.54，验证RMSEv＝3.35 g/kg）相较，精度较高。曹肖奕等[31]优选地面高光谱指数反演土壤电导率，研究表明，土壤电导率的敏感波段位于2000 nm附近；机器学习算法能够显著提高土壤电导率的估算精度。本研究结果与该结论一致。

纹理特征是反演土壤盐分含量的重要依据。研究选用描述影像灰度分布一致性与匀质性的角二阶矩对土壤盐度进行区分[13]。灰度分布均匀时，角二阶矩较大；灰度均匀性弱时，角二阶矩较小。因影像为无植被覆盖的裸土，表面均匀性强的盐土和均匀性弱的非盐碱土存在明显差异，角二阶矩较好区分这种差异。无人机多光谱影像的纹理特征受到多方面的影响，与点对点采集的地面多光谱数据相比，纹理特征不能直观表现出含盐量对土壤的影响程度，而地面高光谱能够直接表现出盐分差异引起的光谱特征变化，使得无人机模型的验证精度低于地面高光谱模型，有待开展对影像纹理特征的深入探究和应用。盐碱土

板结程度不同使无人机影像纹理产生差异，对土壤盐渍化程度的判断具有一定提升。土壤的光谱反射率不仅与盐分含量相关，土壤光谱特征受含水量、有机质和颗粒度等复杂理化性质的影响[32]，在裸土土壤盐度的反演方面仍需要进一步研究。土壤盐度在部分地块差异性较强、变化性大，例如研究区C，对于盐度在15 g/kg以上的样本反演结果偏低，对模型整体精度的影响较大。

5 结 论

本研究基于地面高光谱数据和无人机多光谱数据提出了构建裸土土壤盐度的估测模型。选择统计分析法和机器学习法分别构建PLSR和RF的裸土土壤盐分含量反演模型，并对模型进行验证和优选。获得以下主要结论。

（1）RF盐分反演模型均优于PLSR模型，其中最优的是高光谱RF土壤盐分含量反演模型（R^2=0.54，RMSEv=3.30g/kg）。短波红外优化光谱参量有助于提升估测土壤盐分含量的精度。

（2）无人机影像纹理特征的引入，图谱信息的结合进一步提升土壤盐分含量反演精度，但无人机模型验证精度低于高光谱模型。

（3）黄河三角洲地区土壤盐渍化水平在地理区位上差异明显，南部地区的耕地盐渍化水平较低，适宜种植耐盐碱弱的作物；靠近东北部黄河入海口土壤盐渍化水平较高，宜种植耐盐碱较强的作物。

参考文献：

[1] GORJI T, SERTEL E, TANIK A. Monitoring soil salinity via remote sensing technology under data scarce conditions: A case study from Turkey[J]. Ecological Indicators, 2017, 74: 384-391.

[2] METTERNICHT G I, ZINCK J A. Remote sensing of soil salinity: Potentials and constraints[J]. Remote Sensing of Environment, 2003, 85(1): 1-20.

[3] 王建华, 胡鹏, 龚家国. 实施黄河口大保护 推动黄河流域生态文明建设[J]. 人民黄河, 2019, 41(10): 7-10.
WANG J, HU P, GONG J. Implementing the Great Protection of the Yellow River Estuary and promoting the construction of ecological civilization in the Yellow River basin [J]. People's Yellow River, 2019, 41(10): 7-10.

[4] 刘丽娟, 李小玉. 干旱区土壤盐分积累过程研究进展[J]. 生态学杂志, 2019, 38(3): 891-898.
LIU L, LI X. Research progress on soil salt accumulation process in arid areas[J]. Journal of Ecology, 2019, 38(3): 891-898.

[5] 丁建丽, 姚远, 王飞. 干旱区土壤盐渍化特征空间建模[J]. 生态学报, 2014, 34(16): 4620-4631.
DING J, YAO Y, WANG F. Spatial modeling of soil salinization characteristics in arid areas[J]. Journal of Ecology, 2014, 34(16): 4620-4631.

[6] WANG J, DING J, YU D, et al. Machine learning-based detection of soil salinity in an arid desert region, Northwest China: A comparison between Landsat-8 OLI and Sentinel-2 MSI[J]. Science of the Total Environment, 2020, 707: ID 136092.

[7] 孙亚楠, 李仙岳, 史海滨, 等. 基于多源数据融合的盐分遥感反演与季节差异性研究[J]. 农业机械学报, 2020, 51(06): 169-180.
SUN Y, LI X, SHI H, et al. Remote sensing retrieval of salinity and seasonal differences based on multi-source data fusion[J]. Transactions of the CSAM, 2020, 51(6): 169-180.

[8] HU J, PENG J, ZHOU Y, et al. Quantitative estimation of soil salinity using uav-borne hyperspectral and satellite multispectral images[J]. Remote Sensing, 2019, 11(7): ID 726.

[9] 奚雪, 赵庚星, 高鹏, 等. 基于Sentinel卫星及无人机多光谱的滨海冬小麦种植区土壤盐分反演研究——以黄三角垦利区为例[J]. 中国农业科学, 2020, 53(24): 5005-5016.
XI X, ZHAO G, GAO P, et al. Inversion of soil salinity in coastal winter wheat planting area based on Sentinel satellite and UAV multi spectrum—Taking the reclamation area in the Yellow River Delta as an example[J]. China Agricultural Sciences, 2020, 53(24): 5005-5016.

[10] 贾吉超, 赵庚星, 高明秀, 等. 黄河三角洲典型区域冬小麦播种面积变化与土壤盐分关系研究[J]. 植物营养与肥料学报, 2015, 21(5): 1200-1208.
JIA J, ZHAO G, GAO M, et al. Study on the relationship between the change of winter wheat sowing area and soil salinity in typical regions of the Yellow River Delta[J]. Journal of Plant Nutrition and Fertilizer, 2015, 21(5): 1200-1208.

[11] 黄权中, 徐旭, 吕玲娇, 等. 基于遥感反演河套灌区土壤盐分分布及对作物生长的影响[J]. 农业工程学报, 2018, 34(1): 102-109.
HUANG Q, XU X, LYU L, et al. Inversion of soil salt distribution and its impact on crop growth in Hetao irrigation area based on remote sensing [J]. Transactions of the CSAE, 2018, 34(1): 102-109.

[12] 杨宁, 崔文轩, 张智韬, 等. 无人机多光谱遥感反演不同深度土壤盐分[J]. 农业工程学报, 2020, 36(22): 13-21.
YANG N, CUI W, ZHANG Z, et al. Inversion of soil salinity at different depths by UAV multispectral remote sensing[J]. Transactions of the CSAE, 2020, 36 (22): 13-21.

[13] 黄静, 赵庚星, 奚雪, 等. 光谱与纹理信息结合的黄河三角洲土壤盐渍化信息提取——以垦利区为例[J]. 农业资源与环境学报, 2022, 39(3): 594-601.
HUANG J, ZHAO G, XI X, et al. Extraction of soil salinization information in the Yellow River Delta based on the combination of spectral and texture information —Taking reclamation area as an example[J]. Journal of Agricultural Resources and Environment, 2022, 39(3): 594-601.

[14] 台翔. 植被覆盖条件下的无人机多光谱遥感土壤含盐量监测模型[D]. 杨凌: 西北农林科技大学, 2022.
TAI X. Monitoring model of soil salt content by UAV multi spectral remote sensing under vegetation coverage[D]. Yangling: Northwest Agricultural and Forestry University, 2022.

[15] 高焕君. 衡水湖周边土壤中全盐量、碳酸根、碳酸氢根及氯离子含量的测定[J]. 安徽化工, 2021, 47(5): 96-98.
GAO H. Determination of total salt content, carbonate, bicarbonate and chloride ion content in soil around Hengshui Lake[J]. Anhui Chemical Industry, 2021, 47 (5): 96-98.

[16] ROUSE J W, HAAS JR R H, SCHELL J A, et al. Monitoring vegetation systems in the Great Plains with ERTS[C]// Goddard Space Flight Center 3d ERTS-1 Symphony. Washington DC, USA: NASA, 1974: 309-317.

[17] HUETE A R. A soil-adjusted vegetation index (SAVI) [J]. Remote Sensing of Environment, 1988, 25: 295-309.

[18] GONG P, PU R, G.SBIGING, et al. Estimation of forest leaf area index using vegetation indices derived from Hyperion hyperspectral data[J]. IEEE Transactions on Geoscience and Remote Sensing, 2003, 41: 1355-1362.

[19] CAO Q, MIAO Y, WANG H, et al. Non-destructive estimation of rice plant nitrogen status with crop circle multispectral active canopy sensor[J]. Field Crops Research, 2013, 154: 133-144.

[20] LU J, MIAO Y, SHI W, et al. Evaluating different approaches to non-destructive nitrogen status diagnosis of rice using portable RapidSCAN active canopy sensor[J]. Scientific Reports. 2017, 7: ID 14073.

[21] JORDAN C F. Derivation of leaf-area index from quality of light on the forest floor[J]. Ecology, 1969, 50: 663-666.

[22] REYNIERS M, WALVOORT D J, DE BAARDEMAAKER J. A linear model to predict with a multispectral radiometer the amount of nitrogen in winter wheat[J]. International Journal of Remote Sensing, 2006, 27, 4159-4179.

[23] ALLBED A, KUMAR L, ALDAKHEEL Y Y. Assessing soil salinity using soil salinity and vegetation indices derived from IKONOS high-spatial resolution imageries: Applications in a date palm dominated region[J]. Geoderma, 2014, 230: 1-8.

[24] HARALICK ROBERT M, SHANMUGAM K, ITS'HAK D. Textural features for image classification[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1973, SMC-3(6): 610-621.

[25] 邓远立, 卢伟. 基于灰度共生矩阵的震后倒塌房屋遥感信息提取——以2014年云南鲁甸6.5级地震为例[J]. 华南地震, 2019, 39(2): 100-111.
DENG Y, LU W. Remote sensing information extraction of collapsed houses after earthquakes based on gray level co-occurrence matrix—Taking the 2014 Ludian earthquake of magnitude 6.5 in Yunnan as an example[J]. South China Earthquake, 2019, 39(2): 100-111.

[26] 刘国旗. 多重共线性的产生原因及其诊断处理[J]. 合肥工业大学学报(自然科学版), 2001(4): 607-610.
LIU G. The causes of multicollinearity and its diagnosis and treatment[J]. Journal of Hefei University of Technology (Natural Science Edition), 2001(4): 607-610.

[27] 杨丽萍, 任杰, 王宇, 等. 基于多源遥感数据的居延泽地区土壤盐分估算模型[J]. 农业机械学报, 2022, 53 (11): 226-235.
YANG L, REN J, WANG Y, et al. Estimation model of soil salinity in Juyanze area based on multi-source remote sensing data[J]. Transactions of the CSAM, 2022, 53(11): 226-235.

[28] 王建雯, 李振海, 朱红春, 等. 基于野外实测高光谱数据的盐碱土盐分含量估测研究[J]. 山东科技大学学报(自然科学版), 2017, 36(3): 17-24.
WANG J, LI Z, ZHU H, et al. Research on estimation of salt content in saline alkali soil based on field measured hyperspectral data[J]. Journal of Shandong University of Science and Technology (Natural Science Edition), 2017, 36(3): 17-24.

[29] 翁永玲, 宫鹏. 土壤盐渍化遥感应用研究进展[J]. 地理科学, 2006(3): 369-375.
WENG Y, GONG P. Progress in the application of remote sensing to soil salinization[J]. Geographic Science, 2006(3): 369-375.

[30] 杨丽萍, 任杰, 王宇, 等. 基于多源遥感数据的居延泽地区土壤盐分估算模型[J/OL]. 农业机械学报: 1-14. [2022-11-10]. http://kns. cnki. net/kcms/detail/11.1964. S.20220926.2009.008.html.

YANG L, REN J, WANG Y, et al. Estimation model of soil salinity in Juyanze area based on multi-source remote sensing data[J/OL]. Transactions of the CSAM. 1-14. [2022-11-10]. http://kns. cnki. net/kcms/detail/11.1964.S.20220926.2009.008.html.

[31] 曹肖奕, 丁建丽, 葛翔宇, 等. 基于光谱指数与机器学习算法的土壤电导率估算研究[J]. 土壤学报, 2020, 57(4): 867-877.
CAO X, DING J, GE X, et al. Estimation of soil conductivity based on spectral index and machine learning algorithm[J]. Journal of Soil Science, 2020, 57(4): 867-877.

[32] 郭斌, 白昊睿, 张波, 等. 基于RF和连续小波变换的露天煤矿土壤锌含量高光谱遥感反演[J]. 农业工程学报, 2022, 38(10): 138-147.
GUO B, BAI H, ZHANG B, et al. Hyperspectral remote sensing inversion of soil zinc content in open pit coal mine based on rf and continuous wavelet transform[J]. Transactions of the CSAE, 2022, 38(10): 138-147.

Machine Learning Inversion Model of Soil Salinity in the Yellow River Delta Based on Field Hyperspectral and UAV Multispectral Data

FAN Chengzhi[1], WANG Ziwen[1], YANG Xingchao[1], LUO Yongkai[2], XU Xuexin[3], GUO Bin[1], LI Zhenhai[1*]

(1. *College of Geodesy and Geomatics Information Technology*, *Shandong University of Science and Technology*, *Qingdao* 266590, *China*; 2. *Shandong Key Laboratory of Eco-Environmental Science for Yellow River Delta*, *Binzhou University*, *Binzhou* 256603, *China*; 3. *College of Agronomy*, *Qingdao Agricultural University*, *Qingdao* 266109, *China*)

Abstract: Soil salinization in the Yellow River Delta is a difficult and miscellaneous disease to restrict the development of agricultural economy, and further hinders agricultural production. To explore the retrieval of soil salt content from remote sensing images under the condition of no vegetation coverage, the typical area of the Yellow River Delta was taken as the study area to obtain the hyperspectral of surface features, the multispectral of UAVs and the soil salt content of sample points. Three representative experimental areas with flat terrain and obvious soil salinization characteristics were set up in the study area, and 90 samples were collected in total. By optimizing the sensitive spectral parameters, machine learning algorithms of partial least squares regression (PLSR) and random forest (RF) for inversion of soil salt content were used in the study area. The results showed that: (1) Hyperspectral band of 1972 nm had the highest sensitivity to soil salt content, with correlation r of −0.31. The optimized spectral parameters of shortwave infrared can improve the accuracy of estimating soil salt content. (2) RF model optimized by two different data sources had better stability than PLSR model. RF model performed well in terms of generalization ability and balance error, but it had some over-fitting problems. (3) RF model based on ground feature hyperspectral (R^2=0.54, verified RMSE=3.30 g/kg) was superior to the random forest model based on UAV multispectral (R^2=0.54, verified RMSE=3.35 g/kg). The combination of image texture features improved the estimation accuracy of multispectral model, but the verification accuracy was still lower than that of hyperspectral model. (4) Soil salt content based on UAV multi-spectral imageries and RF model was mapped in the study area. This study demonstrates that the level of soil salinization in the Yellow River Delta region is significantly different in geographical location. The cultivated land in the study area is mainly light and moderate salinized soil with has certain restrictions on crop cultivation. Areas with low soil salt content are suitable for planting crops in low salinity fields, and farmland with high soil salt content is suitable for planting crops with high salinity tolerance. This study constructed and compared the soil salinity inversion models of the Yellow River Delta from two different sources of data, optimized them based on the advantages of each data source, explored the inversion of soil salinity content without vegetation coverage, and can provide a reference for more accurate inversion of land salinization.

Key words: soil salinity; remote sensing; ground feature hyperspectral; UAV multispectral; partial least squares regression; random forest; machine learning